

AIR WAR COLLEGE

AIR UNIVERSITY

CLIMBING THE TOWER OF BABEL:
PERFECTING MACHINE TRANSLATION

by

Aras P. Suziedelis, Lieutenant Colonel, USAF

A Research Report Submitted to the Faculty

In Partial Fulfillment of the Graduation Requirements

16 February 2011

Disclaimer

The views expressed in this academic research paper are those of the author and do not reflect the official policy or position of the U.S. Government or the Department of Defense. In accordance with Air Force Instruction 51-303, it is not copyrighted, but is the property of the United States Government.



Dedication

This research paper is dedicated to those who attempt to communicate with clarity each day, and those who seek to fuse and interpret all forms of communication to better understand meaning.

Additionally, this paper has been made possible by, with, and through the gracious gift of time from professors and experts, good guidance from instructors, mentors, and colleagues, and the caring patience of friends and family. I am ineffably grateful for the broad array of support and encouragement I have received throughout my time researching this most fundamental human activity, and the dazzling technological possibilities of the future. Grazie mille.



Contents

Disclaimer	ii
Dedication	iii
Contents	iv
List of Tables	v
Biography.....	vi
I. Introduction	1
II. History of Machine Translation	4
Earliest Efforts.....	4
US Government Involvement.....	4
Challenges to Perfecting Machine Translation	6
III. Approaches to Machine Translation	7
Artificial Intelligence vs. Intelligence Augmentation	7
Rules-based and Statistics-based.....	8
IV. Current Applications.....	11
The Marketplace.....	11
The Army	12
DARPA	13
Google	14
V. Possibilities 2030-2040	15
Advancements in Technology	15
Advancements in Knowledge and Linguistics	16
VI. Tradespace and Decision Points	20
Tradespace.....	20
Decision Points.....	22
VII. Conclusion.....	25
Leveraging Future Solutions	25
Recommendations	26
Appendix.....	28
Bibliography	31

List of Tables

Table 1. Army Machine Translation Capabilities.....	12
Table 2. Speech Recognition Ambiguity	16
Table 3. Machine Translation Tradespace Factors	20
Table 4. Tradespace Challenges and Potential Solutions	23
Table 5. Recommendations to Achieve Perfect Machine Translation in 2030-2040	25



Biography

Lieutenant Colonel Aras P. Suziedelis is a career US Air Force intelligence officer. He was commissioned in 1990 after four years as an Air Force Reserve Officer Training Corps cadet at the Massachusetts Institute of Technology. He holds a Bachelor's degree in Electrical Engineering from MIT, with additional coursework at Harvard University and at the Colegio de Espana in Spain. Additionally, he has earned Master's degrees in Strategic Intelligence from the National Defense Intelligence College, and in Military Operational Art and Science from the Air Command and Staff College. Lieutenant Colonel Suziedelis has operational, deployed, and staff experience with all intelligence disciplines in Air Force, joint, and coalition environments, and has worked at the Squadron, Group, Center, Major Command, Combatant Command, and Air Staff levels. Furthermore, he spent five years addressing joint information operations and cyberspace personnel, doctrine, and tactics issues, and he is a foreign area officer for Europe.

Prior to attending Air War College, Lieutenant Colonel Suziedelis was the Deputy Group Commander of the 693d Intelligence, Surveillance, and Reconnaissance (ISR) Group at Ramstein AB in Germany. In this position, he was responsible for leading the efforts of 600 personnel generating 24/7 ISR products and services for Predator, Global Hawk, U-2, and other sensitive missions across Europe, Africa, and Asia. Prior to that assignment, he was Commander of the 426th Network Warfare Squadron at Vogelweh, Germany, the only cyber operations squadron in Europe. He led the execution of 24/7 missions to protect and defend military networks from hostile attack. Before his command tour, he led all intelligence efforts for US European Command's counter-terrorism operations, where he created EUCOM's only Intelligence Campaign Plan, and oversaw 105 analysts in the production of intelligence for 11 theater and national war plans.

I. Introduction

The Global War on Terrorism may be the next 100-year war. Because it is a war centered on information and intelligence, we will succeed when we know where, when, and how the enemy intends to strike; we will succeed when all our assets use intelligence cooperatively; we will succeed when our monitoring and robotics technologies have human-level understanding; and we will succeed when we not only get inside the enemy's command cycle but also get into his psyche. Otherwise we fail. Said another way, we will succeed if we take machine intelligence to another level. Otherwise, we fail.¹

Professors Patrick Henry Winston, Sajit Rao,
and Boris Katz, Massachusetts Institute of Technology,
Computer Science and Artificial Intelligence Laboratory, 2010

Language has always been the keystone to understanding. By 2035, technological advancements in machine intelligence and translation hold the promise of dissolving language barriers and finally bringing transparency to this greatest challenge.

The citation above, by Professor Patrick Winston, Director of the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology from 1972 to 1997, describes a logical solution to the existential conflict of our time. The citation encompasses the concepts of knowledge, speed in decision-making, and machines with human intelligence. Fundamentally, it is about understanding.² Taken in a larger view, understanding is relevant not just for our current conflicts, but for all times of conflict and peace, to include humanitarian crises, diplomatic dealings, and routine military operations. In all of these cases, understanding other humans is of paramount importance. As a result, unlocking and deciphering language, and therefore meaning, via perfect machine translation (MT) would be an essential weapon in America's arsenal.

¹ Professors Patrick Henry Winston, Sajit Rao, and Boris Katz, "S3, Taking Machine Intelligence to the Next, Much Higher Level," Working Papers, Cambridge, MA, 27 October 2010, 3.

² Interestingly enough, as military strategists, we see the concept of John Boyd's OODA loop clearly contained within these criteria for victory, as scientists encourage us to get inside the enemy's command cycle and psyche.

Machine translation of language isn't rocket science. It's even more difficult. In the time it has taken to develop nuclear weapons or put man on the moon, we have yet to solve the age-old "language problem." It is a challenge of biblical proportions, as evidenced by the etiology of the Tower of Babel. Having machines translate spoken words with accuracy and precision has been a quest for generations. The question remains, is there a way to return to the time of the Tower of Babel, and have all people understand one another?

Despite leaps in computational power, algorithms, and linguistics, achieving this seemingly straightforward goal remains elusive. European mathematicians patented the first rudimentary translation models in 1932, far before computers brought automation to the problem. In 1954, IBM first harnessed computers to translate Russian into English. At the time, machine translation was promised to be "only five years away."³ In the intervening 60 years, although there has been evolutionary progress, the ability to accurately translate speech from one language into another remains a lingering item on the scientific wish list. In fact, even optimistic futurist Dr. Ray Kurzweil recognizes, "no simple tricks, short of fully mastering the principles of human intelligence, will allow a computerized system to convincingly emulate human conversation."⁴ That said, crucial work continues in this area.

Academic, corporate, and government efforts remain focused on this great human challenge, to communicate unhindered by the lack of a common language. But, there is reason for optimism. The accelerating pace of technology across many fronts appears poised to deliver the right tools and processes to perfect MT. We can imagine a time in the 2030s, when sufficient computational horsepower and sophistication will exist to achieve perfect MT, and the US will

³ Joel Garreau, "Tongue in Cheek," *www.washingtonpost.com*, 24 May 2009. <http://www.washingtonpost.com/wp-dyn/content/article/2009/05/21/AR2009052104697.html>.

⁴ Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (NY: Viking Penguin, 2005). Kindle edition, location 5480-96.

apply this spectacular capability to further a gamut of national security interests.

This research paper describes the potential for developing perfect machine language translation capabilities by 2030-2040. It begins with an overview of historical efforts to harness automation for translation, lists several challenges of translation, and describes methodologies that have been used to address the language problem. Next, the paper depicts a sample of current machine translation capabilities in the military, government, and commercial sectors, and suggests technological advancements we will be able to leverage in the future. Finally, the paper culminates with a discussion of tradespace and decision points for achieving perfect machine translation, and offers benefits we will enjoy with language transparency in the 2030s.



II. History of Machine Translation

Earliest Efforts

Machine translation dates to 1932, when French and Russian scientists secured patents for mechanical multi-lingual dictionaries. At the time, research proliferated in academic circles, and American interest and investment rose throughout the 1940s. Scientists at MIT held the first world-wide MT conference in 1952, followed shortly by IBM's breakthrough efforts.⁵ Subsequently, in the 1950s and 1960s, the governments of the US and USSR aggressively pursued MT capabilities.⁶ As money poured into promising MT experiments in the 1960s, in 1964 the US Government formed the Automatic Language Processing Advisory Committee (ALPAC) to review and assess machine translation progress, as well as the likelihood of future success in computational linguistics and machine translation. The final report was a devastating blow to MT projects underway across the nation.

US Government Involvement

In 1966, the ALPAC concluded, "MT was more costly than human translation and produced inferior results. They saw no hope for improvement in the future."⁷ Given the technology of the time, the report tempered the growing expectation of perfectly automated MT capabilities. As a result, the US Government slashed funding for MT projects,⁸ ceasing nearly

⁵ Dr John Hutchins, *Milestones in the History of Machine Translation*, presented at the Current Issues in Theoretical and Applied Linguistics Conference, Chelyabinsk, Russia, 11 December 2007.

⁶ Dr John Hutchins, *Machine Translation and Computer-Based Translation Aids*, presented at the University of East Anglia, Norwich, UK, January 2003. Especially the CIA and the KGB, the intelligence services of both nations.

⁷ Dale Bostad, *Machine Translation at NASIC: A 50-Year Tradition*, Historical Report, Dayton, OH: National Air and Space Intelligence Center, 2005, 2.

⁸ At the time, 17 institutions were receiving approximately \$25,000,000 for MT R&D efforts.

all the momentum of the moment.⁹ However, as the Cold War ensued, the USAF invested significant resources into MT capabilities such as SYSTRAN. The Foreign Technology Division (antecedent to today's National Air and Space Intelligence Center) used MT tools to translate extraordinary numbers of Russian technical documents.¹⁰ For the Air Force, the manpower and time savings were real and useful. As the success of these early capabilities became widely known, more funding, more projects, and more optimism followed.

In the 1970s and 1980s, DARPA launched a concerted effort to advance speech recognition, machine translation, and the overall concept of natural language processing for military utility. Progress continued in the 1980s and 1990s, as the commercial sector joined efforts to develop MT solutions. By 2002, advancements in MT spurred the US National Institute of Standards and Technology (NIST) to sponsor biannual competitions of MT capabilities and evaluation schema.¹¹ As universities, commercial entities, and government labs developed increasingly sophisticated MT tools, the concurrent need to assess these capabilities arose as well, to determine “how good is MT and how good can it get”?

Over the last 30 years, Artificial Intelligence (AI) techniques have emerged as a novel way to train, rather than merely program, computers to perform complicated tasks such as MT. Adding the promise of human-level intelligence to solve this challenge has buoyed the hope of perfect MT, as the scientific community has transitioned from rules-based MT methods, to corpus-based approaches based on a body of existing language elements and statistical methodologies to align languages. However, despite the evolution of engineering, mathematics,

⁹ Dr Hutchins, 11 December 2007.

¹⁰ Dr Hutchins, January 2003.

¹¹ NIST-led open evaluations such as “OpenMT” provide a test bed for experimentation of MT technologies. Similarly, NIST's Metrics for Machine Translation Challenge, or “MetricsMATR” provides a forum to research and promote innovative techniques that advance the measurement sciences used in MT evaluations.
<http://www.nist.gov/itl/iad/mig/mt.cfm>.

and cognitive approaches to solving the language problem, several inherent challenges exist that complicate what may seem a straightforward technical process.

Challenges to Perfecting Machine Translation

In terms of challenges, first is the scope of the problem--there are over 6,800 distinct languages. Moreover, the best translations are rarely simple word-for-word conversions.¹² Second, machine translation becomes exponentially more difficult, with respect to the length of the phrase. In fact, if an MT algorithm has a database of merely 60,000 words, a sequence of three words could be any of 216 trillion possibilities.¹³ Third, translation occurs on multiple levels—at the lexicon, syntax, and semantics levels—all must be addressed.¹⁴ Fourth, nearly all MT approaches require massive datasets to train the program, on the order of thousands of hours of human-transcribed speech and millions of parallel sentences, as well as data to match language, dialect, slang, and new vocabulary.¹⁵ Finally, a host of other issues must be perfected—decoding algorithms must consider factors such as word alignment, reordering one-to-many translation (compound words in one language may not be compound words in another language), dropped words, and inserted words.¹⁶ Specific challenges and potential solutions that form the tradespace for future MT capabilities are listed in Chapter VI and in the Appendix. Up front, however, to chart a path to perfecting MT, choices must be made regarding the best approach within the context of military relevance.

¹² Dr Raymond E. Slyh, interview by Aras Suziedelis, 22 September 2010.

¹³ Ed Grabianowski, "How Speech Recognition Works," www.HowStuffWorks.com, 2007. <http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition.htm>. Granted, this number is tempered by the rules of a specific language that would disallow specific word orderings, or mandate particular word sequences, (e.g., in English, prepositions can only be followed by articles, etc) but the overall numbers convey the complexity of the machine translation challenge of languages.

¹⁴ Dr John Hutchins, *Uses and Applications of Machine Translation*, presented at Westminster University, London, UK, 20 February 2009.

¹⁵ Dr Raymond E. Slyh, *711 HPW Translation Research and Technology Status*, Dayton, Ohio, September 2010.

¹⁶ Philipp Koehn, *Statistical Machine Translation*, presented at the University of Edinburgh, Edinburgh, UK, 10 September 2007, and, Hutchins, January 2003.

III.

Ap
pro
ach
es

So, this [trillion-word] corpus could serve as the basis of a complete model for certain tasks [in machine translation]—if only we knew how to extract the model from the data.¹⁷

Alon Halevy, Peter Norvig, and Fernando Pereira, Google

to Machine Translation



Artificial Intelligence vs. Intelligence Augmentation

As we move toward the future, consideration must be given to the desired end-state of perfect MT. How much will we expect from revolutionary concepts in Artificial Intelligence (AI), or technological advancements, or human intervention? More to the point, will AI deliver human-level language automation, or will AI fall short and merely deliver intelligence augmentation for MT?

Two fundamental approaches have always described attempts at developing and perfecting machine translation: does one seek total automation, or, does one seek only human-

¹⁷ Alon Halevy, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligence Systems*, (2009): 8.

assisted translation? Some experts believe technology can develop MT capabilities with sufficient speed and precision to produce perfect translations. Others argue machine intelligence will only ever produce translations that are “close enough,” and automated translations will always require human editing. Heretofore, the results of single-pass approaches, or MT that has a single input-output cycle, have been improved upon by adding a second-pass to “clean up” poorly translated sections. In this way, human intervention is certainly a failsafe approach to obtain high-quality translations, but true automation of MT would obviate the inherent weaknesses of human translators and deliver the greatest benefit.

Artificial intelligence has long promised to solve many computational challenges. For decades, we have looked to AI technologies to develop intelligent, human-like machines. Although fantastic work in understanding human cognition and leveraging AI has been achieved at places like MIT since the 1960s, the most renowned scientists in the field, optimistic as ever, cite the need for (and fully expect to see) breakthrough advancements to push AI to the next level of utility.¹⁸ According to MIT Professor Gerald Sussman, a leading AI engineer and innovator for over 35 years, the ability to communicate via language, allowing us to “learn about things we have not experienced and develop a complex cumulative culture,” makes humans unique from all other animals.¹⁹ And so, AI appears to be the exact right technology for enabling and advancing MT efforts.²⁰

Rules-based and Statistics-based

Historically, machine translation has been grounded in two camps: rules-based methods,

¹⁸ Professors Robert Berwick, Gerald Jay Sussman, and Patrick Henry Winston, in separate interviews by Aras Suziedelis, Cambridge, MA, 2-3 November 2010.

¹⁹ Dr Gerald Jay Sussman, “Artificial Intelligence, The Next Steps,” Working Papers, Cambridge, MA, 1 November 2010, 1.

²⁰ Dr Sussman, Working Papers, 1 November 2010, 2-3. According to Professor Sussman, AI will mature rapidly and address many long-standing machine intelligence problems. Moreover, he says it is possible to design and construct self-forming, self-healing, autonomous networks on the fly that are guaranteed to be secure against a range of threats. Such programming techniques would enable both future AI development of human intelligence and future MT capabilities.

and statistics-based methods. More recently, scientists have merged the methods into hybrid solutions to leverage the merits of both.

The earliest MT efforts centered on rules-based algorithms, and resulted in three approaches: Direct Translation, Inter-lingual Translation, and Transfer Translation. In Direct Translation, dictionaries and grammar sets are applied to translate a source language directly into a target language. The result is a word-for-word output, uninformed by syntax and linguistic semantics, that often does not convey the intent of the original language. Inter-lingual Translation leverages a more sophisticated analytical schema to translate words from a source language to a target language via a third intermediary (i.e., Inter-lingua) language. By using a third language (e.g., Esperanto or English, etc), AI processes can better accommodate language semantics.²¹ The third rules-based method, Transfer Translation, harnesses knowledge of the source and target grammar structures, and performs a three-stage series of translations and logic-tree syntheses to achieve a more accurate product. Although this process requires a robust description of source-target language lexicon and structures, it provides the highest fidelity results among rules-based MT.²²

Since the early 1990s, corpus-based statistical MT processes have enabled more comprehensive translation products. There are two types of corpus-based methods: Statistics-based MT and Example-based MT, although both involve significant statistical analysis. These processes exploit statistical patterns inherent in languages. Statistics-based MT selects target language words most likely to correspond to source language words in a given context. It uses mathematical probability to compute word strings and structures. Moreover, Statistics-based MT uses multi-stage processes to sequence target language words in the most appropriate manner,

²¹ Colonel James Koh, *Machine Translation: A Key to Information Supremacy and Knowledge-Based Operations* (Carlisle Barracks: US Army War College, 2003).

²² Dr Hutchins, January 2003.

given some understanding of context. Example-based MT depends upon a database of existing translated phrases and matching algorithms, to relate source language segments to target language segments.²³ Due to language nuances, it requires a massive amount of a priori data and processing power, but can produce highly accurate results. Unfortunately, the need for large databases means statistical methods are less applicable solutions for all the world's languages.

According to Google's principal scientist, Franz Och, "about 100 million words of parallel data are needed in order to build a system that produces reasonably acceptable results. This dependency...limits the number of language pairings available for machine translation."²⁴ For example, there isn't enough richly-detailed data to translate from Greek directly into Thai, so a bridge language such as English must be used, complicating the translation.²⁵ And so, there is no single solution for MT, especially if the required language needs are unknown (as might be the case in future military applications – can we predict where the next crisis will occur?).

Given the large variety of approaches to perform MT, different applications use different methodologies, depending upon available data sets, intended end use, processing power, etc. As might be expected, if there is time and efficiency is not a concern, "the most successful applications are achieved by merging various methods."²⁶

²³ Dr Hutchins, 20 February 2009.

²⁴ Randall Stoss, *Planet Google: One Company's Audacious Plan to Organize Everything We Know*, NY: Simon & Schuster, NY, 2008, 86.

²⁵ Stoss, 86-87.

²⁶ Dr Slyh, interview, 22 September 2010.

IV.

Cu

rre

nt

Ap

The biggest crisis we have in developing world-wide Air Force HUMINT is language.²⁷

Carlos E. Bushman, Lieutenant Colonel, USAF
USAF Human Intelligence Program Manager, 6 January 2011

plications

The Marketplace

These days, the marketplace includes all sorts of personal translation software, handheld translators,²⁸ and voicemail text convertors. Within the DoD, multiple programs are underway to advance MT, but many are narrow and limited to specific contextual dictionaries or languages. This narrowing bounds the problem, but doesn't lead to the most useful solutions. Ultimately, compromises must be made to implement MT. According to Dr Ray Slyh at the Air Force Research Laboratory (AFRL), Wright-Patterson Air Force Base, Dayton, Ohio, "with the competing demands of UAVs and weapons, there isn't interest in the Air Force to fund translators for all 6,800 languages."²⁹ Thus, the reality of resource limitations are a clear

²⁷ Lieutenant Colonel Carlos E. Bushman, "Air Force Human Intelligence," presented to the Maxwell Intelligence Society, Montgomery, Alabama, 6 January 2011.

²⁸ Little more than roughly-linked "8-in-1" dictionaries.

²⁹ Dr Slyh, interview, 22 September 2010.

harbinger that tradeoffs will need to be considered on the path to perfecting machine translation.

US Air Force

The Air Force Research Laboratory (AFRL) leads USAF R&D efforts in speech and language translation applications.³⁰ The programs at AFRL's Speech & Communication, Research, Engineering and Modeling (SCREAM) Laboratory drive the development of MT and a number of inter-related component technologies for a wide range of aerospace applications.³¹ AFRL has several funded efforts, both in-house as well as at the MIT Lincoln Laboratory and other Federally Funded Research and Development Center (FFRDC) organizations, such as

32

these

are

daily.³³

- **GALE** - Global Autonomous Language Exploitation
- **TRANSTAC** - Spoken Language and Communication Translation System For Tactical Use
- **MADCAT** - Multilingual Automatic Document Classification, Analysis and Translation
- **LASER ACTD** - Language and Speech Exploitation Resources (LASER) Advanced Concepts Technology Demonstration (ACTD)

MITRE.

Many of

USAF

programs

in use

The Army

Presciently, the US Army established requirements for a variety of MT capabilities before 9/11. Since 2001, the Army has been developing several combat-related tools, as shown in

Table 1:

Table 1. Army Machine Translation Capabilities

³⁰ More specifically, AFRL's SCREAM Laboratory is advancing speech-to-speech translation and speech-to-text translation applications to meet military requirements.

³¹ SCREAM Lab programs include speech recognition, MT, text-to-speech synthesis, and natural language processing.

³² One such MITRE effort is a project that gives a "voice" to airborne Unmanned Aerial System platforms that would otherwise operate in silence during an in-flight emergency or crisis.

³³ Dr Slyh, Translation Research and Technology Status, September 2010.

In many ways, the Army is at the forefront of defense language requirements. Ground forces understand only too well the need to have the right language at the right time and place. The Army consolidated their MT capabilities into the Sequoyah Foreign Language Translation and the Machine Foreign Language Translation (MFLT)³⁴ programs that are planned to have speech-to-speech translation, speech-to-text translation, and machine translation tools in multiple languages.³⁵ Progress to date has been limited, but future capabilities will provide speech-to-speech translation services for 10 priority languages, and text-to-text tools for 28 languages.³⁶

DARPA

DARPA has several MT capabilities in development, with two major Army programs undergoing field tests. First, GALE has developed speech-to-text translation and MT capabilities for broadcast news and some other text types in Arabic and Mandarin Chinese. The Arabic speech-to-text translation system for news is in daily use in Iraq and at higher echelons. Second, DARPA, partnering with NIST, has fielded a hand-portable TRANSTAC device that executes speech-to-speech translation in the Iraqi dialect of Arabic for boots-on-the-ground soldiers. A Pashtu version has been in field testing since July 2010.³⁷ These efforts, however, do not have the depth of resources available to database managers in the commercial sector.

³⁴ MFLT is the use of a computer to translate text or speech from one language to another. Sequoyah is the Army Program of Record for MT capabilities that are scalable, interoperable, deployable, and available to warfighters at all echelons.

³⁵ Dr Slyh, Translation Research and Technology Status, September 2010.

³⁶ Vaughn Laganosky, *United States Army Machine Foreign Language Translation (MFLT) Requirements and Capability*, presented at the Machine Translation Summit XII, Ottawa, Ontario, Canada, 4 September 2009.

³⁷ Joel Garreau, "Tongue in Cheek," *www.washingtonpost.com*, 24 May 2009. <http://www.washingtonpost.com/wp-dyn/content/article/2009/05/21/AR2009052104697.html>, and, Dr Slyh, *Translation Research and Technology Status*, September 2010.

Google

In the commercial world, many companies have developed MT applications, but perhaps none is better postured for success than Google, the giant of database queries. Google leads the field in MT, as evidenced by Google Translate and by their victory of the 2005 NIST OpenMT challenge. With access to the entire web, the company's "quick rise to the top echelons of the translation business is a reminder of what can happen when Google unleashes its brute-force computing power on complex problems."³⁸ Google's free application translates 52 different languages with generally good results. Google doesn't use the most complex algorithms, but what they lack in formulae, they make up for in sheer size of data. According to their Google Machine Translation Team, "there's no data like more data."³⁹ Google uses United Nations and European Union documents, which are available in six and 23 different languages, respectively. In 2006, Google announced the categorization of one-trillion words of English language collected from the web, far more than any other commercial project.⁴⁰ Google developers used all of these documents to train their translation engine. According to Google's Franz Och, "[t]his technology can make the language barrier go away....it would allow anyone to communicate with anyone else."⁴¹ Google's brute-force approach⁴² has much to recommend it, but since the application is not actually learning the languages, only comparing data pairs, it is unlikely to be the ultimate MT approach. It is rough, but effective. Better ways will be needed to achieve perfect MT.

³⁸ Miguel Helft, "Google's Computing Power Refines Translation Tool," *New York Times*, 8 March 2010. www.nytimes.com/2010/03/09/technology/09translate.html.

³⁹ Alex Franz and Thorsten Branz, "Googleresearch.blogspot.com," *Google Research Blog*, 3 August 2006. <http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>.

⁴⁰ Stoss, 86.

⁴¹ Helft, 2010.

⁴² Professor Patrick Winston at MIT calls this a "bulldozer" approach, as massive amounts of data are moved around as so much dirt (rather than more eloquently analyzed, understood, and translated) .

V.
Poss
ibilit
ies
2030
-

*In the 1960s and 1970s, early work in Artificial Intelligence provided the high challenges that attracted people who put together personal computers, the Ethernet, the ARPANet, bitmap displays, and the forerunners of today's programming languages and programming-language environments. We can expect that history to repeat itself.*⁴³

Professors Winston, Rao, and Katz, MIT,
Computer Science and Artificial Intelligence Lab, 2010

2040



Powerful and sophisticated computational devices undergird the modern machine translation movement. Despite phenomenal increases in computing power, AI, and cognitive science, many challenges must be overcome to translate speech. Ultimately, none of these challenges are insurmountable, and as technology advances, we will achieve perfect MT.

Advancements in Technology

⁴³ Dr Winston et al, 27 October 2010, 15.

Technology is not only advancing, but accelerating in its advancement. As a result, future engineering innovation is not only possible, but inevitable. The predictability of some technological solutions is advocated by noted futurist John Smart. According to Smart, technology itself will drive the need to solve the “language problem.”⁴⁴ Given the development of increasingly sophisticated methods for disseminating and absorbing data,⁴⁵ the proliferation of information production, as well as the diffusion of information, it is inevitable language will cease to hinder the exchange of meaning. As Google TV brings thousands of channels to homes, this number will expand in meaningful ways as programs will be translated into a language of choice, on demand. This is merely the natural next step in an evolution already underway.⁴⁶

The convergence of mobile phones, pagers, personal digital assistants, global positioning system receivers, and data storage devices into an Apple iPhone clearly shows the pace of technological advancement. Moreover, computer processing power has approximately doubled every 18-24 months.⁴⁷ As we reach physical limits to this trend, many expect other technologies to arise (e.g., quantum computing) to allow the trend-line to continue forward. In essence, “we are figuring out how to miniaturize everything that matters to us,” with corresponding language solutions on the horizon.⁴⁸

Advancements in Knowledge and Linguistics

Future efforts in the fields of knowledge and linguistics need to overcome the following classic speech recognition problem, as depicted in Table 2 below:

⁴⁴ Dr John Smart, interview by Aras Suziedelis, 19 October 2010. John Smart is an internationally renowned scholar on accelerating change, and is the founder and president of the Accelerating Studies Foundation.

⁴⁵ Via displays, either wall-mounted, or biologically appended or inserted as we move toward the future.

⁴⁶ Dr Smart, 19 October 2010.

⁴⁷ In accordance with Moore’s famous maxim.

⁴⁸ Dr Smart, 19 October 2010.

Table 2. Speech Recognition Ambiguity

And so, context matters. One promising method to capture communication and determine meaning is the fusion of various sensory data streams, that is, not relying on a single data source. Fusing visual information with existing speech sources can characterize complete activities, or create what researchers call “stories.”⁵⁰ Given that language is more than merely a form of communication, but a way to convey learning, occurrences, experiences, and emotions, the ability to develop truly intelligent machines may rely on the linking and merging of data from the hum

an
sens
es,
cogn

The pronunciation and signal captured from the following very different expressions must be reconciled by a machine translation mechanism:

[horseback rider] and [paperback writer]

To the listening ear, the words *rider* and *writer* are phonetically identical.⁴⁹

ition, and communication. A new focus by AI scientists to characterize a complete environment from more than a single form of communication would be a powerful step forward in the development of intelligent machines.

One such effort to link senses, sentences, and stories (S3) proposed to DARPA by scientists at MIT is called *Genesis*. For now, *Genesis* translates visual information into text (and

⁴⁹ Professor Robert Berwick, interview by Aras Suziedelis, 3 November 2010.

⁵⁰ Dr Winston et al, 27 October 2010, 2-3.

speech) to characterize actions, or rather, to tell a story.⁵¹ If we look forward to the progression of such a unified approach to machine intelligence, this will lead to the ability to capture visual, audio, and other forms of data, then process them to comprehensively describe all aspects of an event, regardless of location or language. In time, this holds great promise in the total, language transparent, characterization of an event.

Are We There Yet?

How will we know when enough is enough to perform perfect MT?⁵² If we place confidence in Kurzweil's well-reasoned projections regarding technological advancement, machines will approach human brain-level intelligence between 2020-2045.⁵³ By 2029 (too early, according to many scientists), Kurzweil expects most discussions to be human-to-machine, as opposed to traditional human-to-human communications. Since computers and networks will be interconnected (like neurons in a brain), once a language is learned, that capability will be proliferated and available across an entire enterprise. Many capabilities can benefit from such advancements, among them machine translation.

Along the way, we will see several generations of ever-improving MT capabilities. DARPA will construct an iPod-sized device to be worn by soldiers, statesmen, or aid workers, with real-time speech translation between languages. Google will develop MT for communications platforms to allow people speaking different languages to understand one another via real-time translation.⁵⁴ As computer processing power increases, and algorithms discern increasingly finer granularity in speech characterization, MT will become a remote sensing tool, with the ability to indicate mental state, emotional condition, even honesty of a

⁵¹ Professor Patrick Winston, interview by Aras Suziedelis, 2 November 2010, and, Dr Winston et al, 27 October 2010, all.

⁵² In terms of raw processing power and mass storage media.

⁵³ Kurzweil, location 5480-96.

⁵⁴ Garreau, 24 May 2009.

speaker.⁵⁵ Small changes in tone, inflection, pitch, or vocabulary will be observed, analyzed, and linked to determine emotion and truthfulness. Such capabilities integrated into other communications devices (video phones, mobile telephones, personal biological communications equipment) would allow more complete (and perhaps unintended) conveyance of language, meaning, and understanding. Ultimately, the question may be, how much “normal communication” will remain, and how much will communication—courtship, love, negotiation, humor, etc—be forever and irreversibly changed?



⁵⁵ Dr Qian Hu, interview by Aras Suziedelis, 24 January 2008.

VI. Tradespace and Decision Points

[W]e, as an Air Force, will continue to put forth tremendous effort to revolutionize our capabilities in language and culture.

[I]f we underestimate the significance of language, region, and culture in our global endeavors, we do so at our own risk, and to the detriment of our effectiveness.⁵⁶

General Norton A. Schwartz, Chief of Staff
United States Air Force, 26 January 2011

Tradespace

As is evident from the preceding chapters, there are a number of approaches to developing, fielding, and employing machine translation capabilities. All together, these criteria define the tradespace for emerging machine translation applications. The series of compromises, tradeoffs, and choices offer incredible flexibility to tailor exactly the right machine translation solution, but clearly depend upon the required end-state.

In many instances, human translators are the right answer to meet translation needs. They can provide a high-level of proficiency, they can assess body language and non-verbal communications, and they can complement translation with personal cultural experience. That said, they can also be the exactly wrong answer to meet translation needs. Human translators can introduce bias, have questionable allegiances, be costly, be the wrong gender for a given environment, etc. Time and again, commanders from the field report they could not trust their human translators. Biased, and perhaps sympathetic to adversaries, human translators may

⁵⁶ Norton A. Schwartz, *Strengthening Air Force Language Skills and Cultural Competencies*, presented at the Department of Defense Language and Cultural Summit, Alexandria, VA, 26 January 2011.

produce misleading, or even life-threatening translation services. There is also the cost to train human translators, one language at a time. Is there a future for human translators?

The Army, for one, believes there will always be a need for human translators.⁵⁷

Language training will still be required for a small range of situations, such as for personal diplomacy, or in other instances where there might be a benefit to show goodwill in physically learning a language. Other than this specialized role, however, automated MT capabilities seem to deliver a significant range of benefits over human translators.

Automated machine translation capabilities are preferable to human translators in many ways essential to military operations. Machine translation gear doesn't get tired, bored, or biased. Moreover, whereas human translators can only be in one place at one time, 100 soldiers could each have their own MT device, and engage in 100 different simultaneous conversations. Although initial costs of development would be high, ultimately, the cost and time savings, as well as gains in flexibility over the individual human training would be enormous (MT does not require healthcare, travel costs, etc). Since automated MT would be a software/hardware solution, almost any need could be met. Table 3 lists a summary of tradespace factors:

- ***Time*** – to develop, train, field, operate
- ***Quality*** – for end-product or services
- ***Ruggedness*** – of operating environment
- ***Cost*** – to develop, train, field, operate, sustain
- ***Openness*** – to constrain or proliferate capability
- ***Security*** – overt or classified development, fielding, and use
- ***Updates*** – universal updates for new terms, or, cross-language queues
- ***Prevalence*** – does every soldier need one; can all soldiers use one capability
- ***Size*** – web application, portable, hand-held

Table 3. Machine Translation Tradespace Factors

⁵⁷ Laganosky, 4 September 2009.

The choice between human translators and automated MT rests in the end use of the capability. As long as human translators are available, with the proper skill, gender, interest, and capacity, they might suffice to meet existing translation needs. If, however, we seek to develop MT that will be better than human translation, and suitable for a wider range of military missions, we need to consider the gamut of challenges and solutions.⁵⁸

Decision Points

A number of tradespace decision points must be considered as we develop future machine translation capabilities. An analysis of major technical obstacles to perfecting machine translation follows in Table 4. Furthermore, the Appendix contains a more detailed description of tradeoffs and solutions, shaped by military context.

USAF Opportunities

In the 2030-2040 timeframe, the ubiquitous ability to translate and understand languages in real-time will reap many benefits. From data collection to situational awareness, from force protection to coalition dynamics—perfect MT will be a massive force multiplier. From an intelligence collection perspective, MT will allow us to filter vast amounts of raw information obtained by prolific future ISR sensors.⁵⁹ Given continued US supremacy in crypto-analysis, most if not all adversary communications will be vulnerable to exploitation, providing

⁵⁸ According to Garreau's discussion with MT principal investigators at DARPA, there are a number of scales that measure accuracy of machine translation output (e.g., the Metric for Evaluation of Translation with Explicit Ordering [METEOR], and Bilingual Evaluation Understudy [BLEU], are two such common evaluation tools). One such evaluation tool is based upon scoring output on a 1-to-100 point scale: a score of 40 is sufficient to relate the general gist of a translated language; Google's massive database approach results in translations with scores in the 50s; and, 70 is as good as the best human translators can manage. DARPA has developed an MT device that can translate, within a narrow contextual field, with an output score of 64. A human translator conducting the same assessment bested the MT device by only 10-points, scoring 74. Thus, MT is already approaching human translation capabilities for select languages.

⁵⁹ Lieutenant General David Deptula (Retired), former USAF Deputy Chief of Staff for ISR, presentation to the Maxwell Intelligence Society, 14 February 2011, used with permission. According to Lt Gen Deptula, given the increasing surge of ISR full motion video data, and lack of requisite human resources to exploit and analyze massive quantities of data, "We can't afford to have analysts sitting and staring at a screen, waiting for someone to exit a building. We need automation and computers to make it [processing, exploitation, and dissemination] work." Such a statement by one of the top ISR experts in the DoD is a clear mandate for the military utility of MT, especially multi-discipline approaches such as *Genesis*.

exceptional indications and warning (I&W) capabilities and predictive battle-space awareness.

Tradespace Challenges and Potential Solutions

Challenge: Too complex a problem set. There are 6,800 spoken languages on Earth.

Potential Solutions

- (1) Pare down to a subset.
- (2) Group languages by type.
- (3) Allow the private sector to develop MT for common languages.
- (4) Model all languages as variants of a single universal (core) language.

Challenge: Ambiguity. “The same meaning can be expressed in many different ways, and the same expression can express many different meanings.”⁶⁰ A single language includes so many variables, it can be impossible to discern meaning: slang, dialects, tones, and inflections (common in Arabic and Chinese) complicate MT.⁶¹

Potential Solutions

- (1) Collect more baseline data.
- (2) Leverage methodologies that consider context.

Challenge: Clarity of input. “Input data is invariably noisy and incomplete.”⁶² Speech converted to text is degraded via ambient noise levels, compression noise, etc.

Potential Solutions

- (1) Reduce ambient noise.
- (2) Filter the source signal.
- (3) Disentangle discrete conversations.

Challenge: Data sparseness. There is insufficient data to completely and without error account for all language variables—colloquial expressions, connotation, acronyms, etc.

Potential Solutions

- (1) Narrow the scope.
- (2) Compile a larger database.

Table 4. Tradespace Challenges and Potential Solutions

Moreover, instantaneous machine translation will benefit the analytical community, by allowing cross-lingual analysis via network chat in coalition environments. This will be especially useful

⁶⁰ Halevy et al, 11.

⁶¹ Dr Douglas Jones, interview by Aras Suziedelis, 20 October 2010. According to Dr Jones from the MIT Lincoln Laboratory, “the fundamental challenge of human language processing remains ambiguity.”

⁶² Dr Sussman, Working Papers, 2.

for the USAF, as the majority of military ISR analysis and exploitation (e.g., especially from UAS platforms) is conducted by Airmen. Additionally, MT will allow us to search the world for the best educational and training opportunities, identify world-class sources, and make them our own, despite differences in language.

In terms of situational awareness, watch centers and headquarters will be able to tailor pertinent data streams to build maximum situational awareness. Perfect MT will enable the mining of all languages and all mediums for real-time crisis monitoring, characterization of forces, and understanding of global public opinion.

Force Protection will be significantly enhanced by unprecedented MT in all languages. No longer will personnel attempting to gain illicit entry into US facilities, or into the US proper, be able to hide behind barriers of language. On-demand perfect MT will ensure our military and homeland security forces will be able to converse with and question personnel in all languages.

Coalition operations will benefit from previously unimagined levels of understanding and cooperation. Perfect MT will enable unprecedented opportunities to save US resources, as it will enable mission execution by, with, and through like-minded international partners. MT, by virtue of establishing shared understanding, promises to deliver far fewer cultural misunderstandings, language barriers, and tactical/operational/strategic policy errors.

Given the opportunities afforded by perfect MT, the USAF should start planning now (given POM cycles) to leverage this game-changing emerging toolkit.

VII

•
Co
ncl
usi

Our country needs to use our increasing understanding of human intelligence to take machine intelligence to another level to maintain and strengthen our computational economic and military advantage.⁶³

Professors Winston, Rao, and Katz, MIT,
Computer Science and Artificial Intelligence Lab, 2010

on

63



Leveraging Future Solutions

As intractable as the language challenge appears to be, the inevitability of technology will bear fruitful solutions. Accelerating advances in computer technology, to include AI, the convergence of information-based devices, and the nearly-there maturity of MT will combine to realize perfect MT within the next 30 years. Scientists, engineers, and linguists see their fields of study poised and ready for a new round of advancements. Accordingly, the USAF should begin planning to understand and integrate this powerful future capability.

The USAF should consider a range of "organize, train, and equip" changes in anticipation of MT advancements. First, in terms of organization, significant manpower savings could be

⁶³ Dr Winston et al, 27 October 2010, 1.

achieved via perfect MT.⁶⁴ How many translators or crypto-linguists would be needed, if MT could accomplish the same tasks better, faster, and cheaper? How much savings could be garnered by eliminating single language training? Second, perfect MT would have significant impact on all types of USAF training and education. The USAF should plan on using MT in the 2030s to locate, translate, and implement the best educational programs, tools, and experts in the world, regardless of original language or medium. All knowledge would be accessible via an individually tailored/preferred language mix. Finally, perfect MT could impact all USAF operations, by more precisely and properly equipping our forces for warfighting, peacekeeping, humanitarian operations, etc. Perfect MT would permit instantaneous availability of all languages. We would engender closer physical and functional cooperation with partner nations. We would facilitate dialogue between US forces and combatants/civilians in crisis situations across the globe. We would eliminate the barriers to understanding in challenging combined environments. In short, communication and understanding could be guaranteed.

Recommendations

There are no commonly recognized “next steps” to achieve perfect MT. In lieu of the exact answer, the following suggestions outlined in Table 5 will help us progress:

- Continue to develop the next steps of AI to build an intelligent machine. The same innovators that led us to current achievements in AI are aware of the weaknesses to the current approaches, as well as the methods to remedy them. Focus on new flexible programming technology, to accommodate change, project into the future, and establish self-forming networks with inherent security.
- Use machine intelligence to fuse and analyze data from all five human senses (as a human would), but with highly technical analytical capabilities. Determine the relationships between the senses, time, and space, to determine a complete understanding of an environment or activity.
- Develop a single core language, from which all other languages can be projected or determined. By studying this core language (and corresponding deltas that describe other languages), we can achieve a toolkit for all current and future languages.⁶⁵

Table 5. Recommendations to Achieve Perfect Machine Translation in 2030-2040

Perfect MT is a long-standing scientific pursuit. The rapid advancement of technology within the next 20 years will allow us to create perfect MT to leverage the awesome power of instantaneous translation. Ultimately, we will learn how to use new forms of content, new concepts of connectivity, and new methods of understanding, to address the full-range of national security challenges.



⁶⁵ This would allow us, conceivably, to interpret any form of spoken communication, or create a universal translator of sorts, as was featured in the Star Trek television series and movies. Alternatively, if automated machine translation emerges as a highly functional, yet imperfect capability, we may recreate the opportunity for “Navajo Code Talkers” to arise, as low-density or rare languages may not be discernable by near-perfect (still imperfect) MT. Just as in World War II, the use of an obscure verbal form of communication might allow the secret passing of information.

Appendix

Challenge: Too large and complex a problem set.

Linguists generally acknowledge there are approximately 6,800 spoken languages on Earth. This sheer number of languages in itself is daunting, and underscores the complexity of the MT problem.

Potential solutions:

- **(1) Pare down to a smaller subset of militarily, diplomatically, or economically relevant languages.** This is a potential solution, in that it would narrow the scope of the problem set. For example, some defense planners indicate we are only interested in approximately 100 or so languages at all. Moreover, we may expect to be engaged in a still narrower subset of regions, such that our concerns are further limited to just 20-40 languages. That said, history shows it is difficult to determine what smaller subset of languages might be needed to meet future military, diplomatic, or economic requirements. Future calamity, crisis, and natural disaster are impossible to predict with certainty. Moreover, the need to act swiftly and with precision may require a small-scale, but exquisitely equipped force on short-notice, or no-notice. This drives a related, but different requirement of speed—how quickly could an MT capability be developed and fielded in sufficient quantity (and quality) to meet an emerging or imminent need?
- **(2) Group languages by type, given sufficient commonality.** Many languages of the world share some degree of commonality. For example, Spanish, Portuguese, French, Catalan, Sardinian, and Romanian all belong to the Italic sub-group of the Indo-European language family, and share a common linguistic construct. Collecting linguistically related languages and attacking MT challenges by group holds the promise of efficiency savings and wider application of solution sets.
- **(3) Allow the private sector to develop effective MT for commercially-viable languages.** Current efforts by profit-seeking commercial companies (e.g., SYSTRAN, SDL Language Weaver, AppTek, etc) may meet some US Government military requirements for MT, but will not meet all the needs for all required languages. Although common (major or profitable) languages are being addressed now to some extent, and others will likely be candidates in the future, there will undoubtedly remain many low-density languages or precision needs that will not be met by commercial products.
- **(4) Model all language as variants of a single universal (core) language of rules and semantics, and solve MT for this single common core language, plus variation deltas.** Although a preliminary proposal, it may be worthwhile to examine the ability to determine the essential core or universal components of language en toto, then calculate

the differences needed to fashion the core language into all existing languages. This experimental approach might considerably shorten the time it would take to add a new language to an existing automated machine translation capability.

Challenge: Ambiguity of words and meanings.

Ambiguity is among the most classically perplexing challenges to perfect machine translation. “The same meaning can be expressed in many different ways, and the same expression can express many different meanings.”⁶⁶ According to Massachusetts Institute of Technology Lincoln Laboratory researchers focused on MT, “the fundamental challenge of human language processing remains ambiguity.”⁶⁷ A single language includes so many variables, it can be impossible to uniquely discern and characterize sounds of speech, and therefore meaning. In addition, spoken speech lacks audible clarity, even with wildly different meanings. Recall the words *rider* and *writer* are pronounced and sound exactly the same. Moreover, slang terms, dialects, and even tonal variations and inflections (such as is common in Arabic languages and Chinese dialects) further complicate MT.

Potential solutions:

- **(1) Collect more baseline data.** Further progress can be made via greater collection and analysis of massive amounts of data. As digital storage and computer processing power costs continue to drop, leverage these technologies and open access to the web to collect rich sources of multi-lingual data: “the models become better and better the more text you process.”⁶⁸
- **(2) Leverage methodologies that consider context of words.** Consider context, in addition to rules, if time and storage space is available in the solution mechanism. In the future, processing power and magnetic storage capacity technology will clearly allow great efficiencies for MT.

Challenge: Clarity of input.

In many cases, MT is hindered by a lack of clarity of signal, or input data. “Input data is invariably noisy and incomplete.”⁶⁹ If a target text is based on speech that has been received for speech-to-text translation, and is then later used for MT, there may be significant degradation in the original source language: e.g., high ambient noise levels, compression or modulation noise, tonal changes, etc. Impurities in the original capture of voice would induce significant challenges in the speech-to-text, and subsequent MT processes.

⁶⁶ Halevy et al, 11.

⁶⁷ Dr Douglas Jones, interview by Aras Suziedelis, 20 October 2010.

⁶⁸ Helft, 2010. Comment by Dr Franz Josef Och, principal scientist and lead for machine translation at Google.

⁶⁹ Dr Sussman, Working Papers, 2.

Potential solutions:

- **(1) Reduce ambient noise.** There are a number of approaches to capture audio for speech-to-text translation to feed subsequent MT efforts. This is an especially significant problem in military environments, where there can be a range of background noise and vibration (cacophony of front lines, battle rattle, aircraft, vehicles, etc). From a hardware perspective, MITRE has recently conducted tests of non-traditional microphone devices, in an attempt to isolate voice signals from background noise. To date, ear and neck microphones appear to show great promise to capture original source voice with higher degrees of purity.⁷⁰
- **(2) Filter and clean the source signal.** From a software or computational view, input data can be processed and cleaned. “[T]here are mechanisms that heuristically complete incomplete data, clean noise from data, and generally make sense of unfamiliar situations.”⁷¹
- **(3) Disentangle discrete conversations.** In a typical room of 10 people, there are on average, 1.76 multiple simultaneous conversations. In a chat room, there are even more multiple simultaneous conversations, an average of 2.7 among ten participants.⁷² Mathematical models exist, and can be improved upon, to untangle these simultaneous communication threads, to make sense of what may appear to be non-sequitor sequences of words and meanings.

Challenge: Data sparseness.

There exists an insufficient set of data to completely and without error account for all language variables—colloquial expressions, word sequences, connotation, acronyms, etc.

Potential Solution:

- **(1) Narrow the scope of the effort.** Consider narrowing the scope of the MT context then fully characterize and populate this smaller language space (treaty negotiation terminology, tactical military expressions, strategic diplomatic issues, etc).
- **(2) Compile a large database.** Include all possible cases. Google is working toward this end, via their peerless access to the totality of data on the web. Google has amassed a database of over two trillion matched words, and has made the database public to spur academic, commercial, or government MT success.

⁷⁰ Dr Qian Hu and Steve Jones, interview by Aras Suziedelis, 1 November 2010.

⁷¹ Dr Sussman, Working Papers, 2.

⁷² Micha Charniak and Eugene Elsner, "Disentangling Chat," *Computational Linguistics* 36, no. 3, September 2010, 390.

Bibliography

Berwick, Robert, interview by Aras Suziedelis. Cambridge, MA, (3 November 2010).

Bostad, Dale. *Machine Translation at NASIC: A 50-Year Tradition*. Historical Report, Dayton, OH: National Air and Space Intelligence Center, 2005, 2.

Brants, Alex Franz and Thorsten. "Googleresearch.blogspot.com." *Google Research Blog*. 3 August 2006. <http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>.

Bushman, Carlos E. *Air Force Human Intelligence*. Presented to the Maxwell Intelligence Society at Maxwell Air Force Base, Montgomery, Alabama, 6 January 2011.

Charniak, Micha Elsner and Eugene. "Disentangling Chat." *Computational Linguistics* 36, no. 3 (September 2010): 389-409.

Condon, Sherri, interview by Aras Suziedelis. Bedford, MA, (1 November 2010).

Deptula, David. Presentation to the Maxwell Intelligence Society at Maxwell Air Force Base, Montgomery, Alabama, 14 February 2011. Used with permission.

Deptula, David. Presentation to the Air War College at Maxwell Air Force Base, Montgomery, Alabama, 15 February, 2011. Used with permission.

Garreau, Joel. "Tongue in Cheek." *washingtonpost.com*. 24 May 2009. <http://www.washingtonpost.com/wp-dyn/content/article/2009/05/21/AR2009052104697.html> (accessed 26 September 2010).

Grabianowski, Ed. "How Speech Recognition Works." *HowStuffWorks*. 2007. <http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition.htm> (accessed 28 September 2010).

Halevy, Alon, Peter Norvig, and Fernando Pereira. "The Unreasonable Effectiveness of Data." *IEEE Intelligence Systems*, 2009: 8-12.

Helft, Miguel. "Google's Computing Power Refines Translation Tool." *New York Times*. 8 March 2010. www.nytimes.com/2010/03/09/technology/09translate.html.

Hu, Qian, interview by Aras Suziedelis. Bedford, MA, (1 November 2010).

Hu, Qian, interview by Aras Suziedelis. Bedford, MA, (24 January 2008).

Hutchins, John. *Machine Translation and Computer-Based Translation Aids*. Presented to the University of East Anglia. Norwich, January 2003.

Hutchins, John. *Machine Translation: Past Imperfect, Future Indefinite*. Presentation at the University of Leeds. Leeds, November 2003.

Hutchins, John. *Milestones in the History of Machine Translation*. Presented at the Current Issues in Theoretical and Applied Linguistics Conference. Chelyabinsk, 11 December 2007.

Hutchins, John. *Uses and Applications of Machine Translation*. Presentation to Westminster University. London, 20 February 2009.

Jones, Douglas, interview by Aras Suziedelis, (20 October 2010).

Jones, Stephen, interview by Aras Suziedelis. Bedford, MA, (1 November 2010).

Jones, Steven, interview by Aras Suziedelis. Bedford, MA, (24 January 2008).

Koehn, Philipp. *Statistical Machine Translation*. Edinburgh, 10 September 2007.

Koh, James. *Machine Translation: A Key to Information Supremacy and Knowledge-Based Operations*. Carlisle Barracks: US Army War College, 2003, 40.

Kuehl, Daniel T. "From Cyberspace to Cyberpower: Defining the Problem." In *Cyberpower and National Security*, edited by Stuart H. Starr, Larry K. Wentz Franklin D. Kramer, 40. Washington, DC: National Defense University Press, 2009.

Kurzweil, Ray. *The Singularity is Near: When Humans Transcend Biology*. NY: Viking Penguin. Kindle edition, Location 5480-96., 2005.

Laganosky, Vaughn. *United States Army Machine Foreign Language Translation (MFLT) Requirements and Capability*. Presented at the Machine Translation Summit XII, Ottawa, Ontario, Canada, 4 September 2009.

Patrick Henry Winston, Sajit Rao, Boris Katz. "S3, Taking Machine Intelligence to the Next, Much Higher Level." Working Papers. Cambridge, MA, 27 October 2010. 17.

Schwartz, Norton. *Strengthening Air Force Language Skills and Cultural Competencies*. Presented to the Department of Defense Language and Culture Summit., 26 January 2011.

Slyh, Raymond E., interview by Aras Suziedelis. Dayton, OH, (22 September 2010).

Slyh, Raymond E. *711 HPW Translation Research and Technology Status*. Dayton, Ohio, 20 September 2010.

Smart, John, interview by Aras Suziedelis. Montgomery, AL, (19 October 2010).

Stoss, Randall. *Planet Google: One Company's Audacious Plan to Organize Everything We Know* 2008. Ny: Simon & Schuster, NY.

Sussman, Gerald Jay, interview by Aras Suziedelis. Cambridge, MA, (2 November 2010).

—. "Artificial Intelligence, The Next Steps." Working Papers. Cambridge, MA, 1 November 2010.

Winston, Patrick Henry, interview by Aras Suziedelis. Cambridge, MA, (2 November 2010).

